

A Lower Bound on the Complexity of Approximating the Entropy of a Markov Source

Travis Gagie

Department of Computer Science
University of Chile
travis.gagie@gmail.com

The Asymptotic Equipartition Property (see, e.g., [3]) implies that, if we choose the characters of a string s of length n independently and according to the same probability distribution P over the alphabet then, for large values of n , the 0th-order empirical entropy $H_0(s)$ of s (see, e.g., [4]) will almost certainly be close to the entropy $H(P)$ of P . Batu, Dasgupta, Kumar and Rubinfeld [1] showed that, if $H(P) = \Omega(\gamma/\epsilon)$, then we can almost certainly approximate $H(P)$ to within a factor of γ after seeing $\mathcal{O}(\sigma^{(1+\epsilon)/\gamma^2} \log \sigma)$ characters of s , where σ is the alphabet size and ϵ is any positive constant; they proved a lower bound of $\Omega(\sigma^{1/(2\gamma^2)})$, which was later improved by Raskhodnikova, Ron, Shpilka and Smith [5] and Valiant [6].

Similarly, the Shannon-McMillan-Breiman Theorem (see, e.g., [3] again) implies that, if we generate s from a stationary ergodic k th-order Markov source \mathcal{X} then, for large values of n , the k th-order empirical entropy $H_k(s)$ of s (see, e.g., [4] again) will almost certainly be close to the entropy $H(\mathcal{X})$ of \mathcal{X} . Although many papers have been written about approximating the entropy of a Markov source based on a sample (see, e.g., [2] and references therein), we know of no upper or lower bounds similar to Batu et al.'s results. We now give a simple proof that, even if we know \mathcal{X} has entropy either 0 or at least $\log(\sigma - k)$, there is still no algorithm that, with probability bounded away from $1/2$, guesses its entropy correctly after seeing at most $(\sigma - k)^{k/2-\epsilon}$ characters.

Lemma 1. *For any $k \geq 1$, $\epsilon > 0$ and sufficiently large σ , there is a k th-order Markov source over the alphabet $\{0, \dots, \sigma - 1\}$ that has entropy at least $\log(\sigma - k)$ but, with high probability, does not emit duplicate k -tuples among its first $(\sigma - k)^{k/2-\epsilon}$ characters.*

Proof. Consider the k th-order Markov source that, whenever it has emitted a k -tuple $\alpha = a_1, \dots, a_k$, emits a character drawn uniformly at random from $\{0, \dots, \sigma - 1\} - \{a_1, \dots, a_k\}$. Notice this source has entropy at least $\log(\sigma - k)$. Also, a k -tuple α cannot occur in position i if it occurs in any of the positions $i - k + 1, \dots, i - 1, i + 1, \dots, i + k - 1$, and vice versa. Finally, the probability α occurs in position i is independent of whether it occurs in position j for $j \leq i - k$ or $j \geq i + k$.

For $i - k + 1 \leq j \leq i + k - 1$, let the indicator variable B_j be 1 if α occurs in position j , and 0 otherwise. By Bayes' Rule, the probability α occurs in position i , given that it does not occur in any of the positions $i - k + 1, \dots, i - 1, i + 1, \dots, i + k - 1$, is

$$\begin{aligned} & \Pr[B_i = 1 \mid B_{i-k+1} = \dots = B_{i-1} = B_{i+1} = \dots = B_{i+k-1} = 0] \\ &= \frac{\Pr[B_i = 1 \text{ and } B_{i-k+1} = \dots = B_{i-1} = B_{i+1} = \dots = B_{i+k-1} = 0]}{\Pr[B_{i-k+1} = \dots = B_{i-1} = B_{i+1} = \dots = B_{i+k-1} = 0]} \\ &\leq \frac{\Pr[B_i = 1]}{1 - \Pr[B_{i-k+1} = 1 \text{ or } \dots \text{ or } B_{i-1} = 1 \text{ or } B_{i+1} = 1 \text{ or } \dots \text{ or } B_{i+k-1} = 1]} \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1/(\sigma - k)^k}{1 - (2k - 2)/(\sigma - k)^k} \\
&= \frac{1}{(\sigma - k)^k - 2k - 2}.
\end{aligned}$$

It follows that the probability α occurs at least twice among the first $(\sigma - k)^{k/2-\epsilon}$ emitted characters is at most the probability that, while drawing $(\sigma - k)^{k/2-\epsilon}$ elements uniformly at random and with replacement from a set of size $(\sigma - k)^k$, we draw a specified element at least twice. Therefore, the probability any k -tuple occurs at least twice among the first $(\sigma - k)^{k/2-\epsilon}$ emitted characters is at most the probability that we draw any element at least twice. For $k \geq 1$ and sufficiently large σ , both probabilities are negligible. \square

Theorem 1. *Suppose that, for any $k \geq 1$, $\epsilon > 0$ and sufficiently large σ , we are given a black box that allows us to sample characters from a k th-order Markov source over the alphabet $\{0, \dots, \sigma - 1\}$. Even if we know the source has entropy either 0 or at least $\log(\sigma - k)$, there is still no algorithm that, with probability bounded away from $1/2$, guesses the entropy correctly after sampling at most $(\sigma - k)^{k/2-\epsilon}$ characters.*

Proof. Consider any algorithm A for guessing the source's entropy. Suppose there is a string s of length $(\sigma - k)^{k/2-\epsilon}$ containing no duplicate k -tuples and such that, with probability at least $1/2$, A stops and guesses “at least $\log(\sigma - k)$ ” after sampling a prefix of s . Then on any source with entropy 0 that starts by emitting s with probability 1 the algorithm errs with probability at least $1/2$. Given s , it is straightforward to build such a source.

Now suppose there is no such string s . Then whenever the first $(\sigma - k)^{k/2-\epsilon}$ sampled characters contain no duplicate k -tuples, A either samples more characters or stops and guesses “0”, with probability at least $1/2$. Therefore, on any source with entropy at least $\log(\sigma - k)$ that, with high probability, does not emit duplicate k -tuples among its first $(\sigma - k)^{k/2-\epsilon}$ characters — such as the one described in the lemma above — A either samples more characters or errs, with probability nearly $1/2$. \square

References

1. T. Batu, S. Dasgupta, R. Kumar, and R. Rubinfeld. The complexity of approximating the entropy. *SIAM Journal on Computing*, 35(1):132–150, 2005.
2. H. Cai, S. R. Kulkarni, and S. Verdú. Universal entropy estimation via block sorting. *IEEE Transactions on Information Theory*, 50(7):1551–1561, 2004.
3. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.
4. G. Manzini. An analysis of the Burrows-Wheeler transform. *Journal of the ACM*, 48(3):407–430, 2001.
5. S. Raskhodnikova, D. Ron, A. Shpilka, and A. Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. In *Proceedings of the 48th Symposium on Foundations of Computer Science*, pages 559–569, 2007.
6. P. Valiant. Testing symmetric properties of distributions. In *Proceedings of the 40th Symposium on Theory of Computing*, pages 383–392, 2008.